TerraWeather
Ensuring Your Safety & Success

AI Workshop – Outline

An Intuitive Approach to Deep Learning for Time-series Forecasting

0.0 Goals

This 6-day course aims to teach you:

- **Neural Network (NN) techniques for timeseries forecasting**. We focus on neural networks only, using the Caffe [1] deep learning framework. We will only cover timeseries forecasting, but the techniques you learn should be applicable to classification problems.

- **Best practices for NN based forecasting**. How to determine optimum network configurations? How to train complex networks? How to evaluate and track performance? When to apply a technique and when not to?

- **An appreciation of NNs from a mathematical perspective.** You have to understand the math to be able to design your own systems. We will take a bottom-up approach to teaching you how to translate the math into practical rules of thumb you can use to design and train NNs.

In short, this course will help you design effective neural networks for solving real-world problems.

0.1 Course Format

The lectures and labs will be backed by a real-world problem: predicting the outbreak of dengue [2] using environmental data.

Each Day of the course has two 2-hour sessions, each comprising a Lecture followed by a Lab. (ie, 2 lectures and 2 labs a day). You must attend all sessions. Each Day ends with a group Homework that your group should complete before the next Day's sessions. We will call groups to present their work at the start of each Day.

0.2 Prerequisites

1. **You should already know a programming language** (eg, Python). Because we are using Caffe, this is not a strict requirement, but it will be very helpful for the Datathon and to complete some Labs. At the minimum, you should ensure that someone in your Group is a capable programmer.

2. **You should be comfortable with basic multivariate Calculus**

(differentiation, integration over multiple variables). You need this background to appreciate the mathematical underpinnings of NNs.

3. **You need a decent laptop.** All lab sessions will be done on your laptop using Virtual Box[3]. Your laptop needs to be: 64-bit, have at least 10GB of free disk space and 8GB of RAM. Bring your laptop fully charged and charger to all workshop sessions.

4. **Install VirtualBox.** Your OS must be capable of running Virtual Box. Windows, MacOS and Linux should be OK. Please download and install Virtual Box version 5.2.0 before the course commences.
(**If you run Linux** on your laptop, you should not use Virtual Box, but instead download and compile Caffe 1.0 directly onto your laptop. Do **not** use apt-get or yum. Compile from source, because we will patch Caffe with Terra Weather modifications for timeseries.)

0.3 Course Outline

Here is a tentative outline of the course. This is still subject to change.

DAY 1
Lecture 1: Introduction to Caffe & Linux • Brief overview popular frameworks. GPU & CPU. CuDNN, MXNET; Keras/Theano/Tensorflow. • Data exploration tools (Overture / Python Notebook / Jupiter Notebook) • What is Caffe. Why we use it. Why it is good to learn it. • Terra Weather's additions to standard Caffe / Autocaffe, for automating Caffe. **Lab 1:** Using VirtualBox. / Basic BASH commands. cd, ls, vim, ssh, scp, cp, mv, rm, chmod running shell scripts. Simple hello world in BASH / Getting Jupiter Notebook to work. / Transfer files into/out of Linux. **Lecture 2: Running Caffe** • Caffe overview of mechanics: Prototext file format, solver file, resources. • Caffe's dimensions. • Creating new input data using Python & HDF5. • How to examine & visualize the output? • Introduce autocaffe functions for scheduling work. **Lab 2:** Running toy example: Sine wave prediction / Double sine prediction.

Homework: Construct Median, Min, Max of weekly temperature & rainfall data into HDF5 format from raw datasets. This homework requires Python programming background.

DAY 2

Lecture 3: Neural Networks Basics

- What they are (basic building blocks) / Why they work. / Interpretation of output as class, probability or value.
- Simple networks (MLP) with simple loss layer.
- Backpropagation Algorithm / Mathematical Underpinnings
- Caffe's forward & backward passes.
- Training / Testing

Lab 3: Predicting the median temperature 1, 4, 8 weeks ahead.

Lecture 4: Training and Evaluating Neural Networks

- How good is good? Using benchmarks to evaluate performance.
- Need to collect statistics by re-running experiments.
- Mini-batches.
- Overfitting: What it is / Detecting / Early Stopping / Regularization / Dropouts.
- Using AutoCaffe for simple visualization with Latex templates.

Lab 4: Predicting the median temperature 1, 4, 8 weeks ahead ("time horizon") with persistence benchmark. Try predicting rainfall (optional)

Homework: The dengue dataset. Simple networks to predict dengue. Compare predictions with Persistence. Students to try 1, 8 & 16 week predictions. Groups need to come up with a summary of their findings. Selected groups will present a 3 - 5 min the next Day, with 5 min Q&A each.

DAY 3

Lecture 5: Timeseries Data

- What sets apart timeseries data?
- NNs are not black boxes / need for preprocessing & good loss functions.
- Basic Input Preprocessing: Input Normalization / Why it is important.
- Eliminating Dependent input variables / Why it is important

Lab 5: Autocaffe "expander" system for automating different configurations.

Lecture 6: The Loss Function

- Importance of Loss function / Why naive losses are not sufficient for timeseries predictions.
- Loss functions: Euclidean / Momentum / Force / How these work
- Evaluating performance: Autocorrelation & Lag / Why lag itself is not sufficient.

Homework: Implement Input scaling & Momentum & Force losses for dengue predictor. Same time horizons. Students need to get stats of their results, and report their best scores on a special forum chat. We will select groups to present their results the next Day. (10 mins + 10 mins Q&A). All groups should prepare their presentation.

DAY 4

Lecture 7: Deep Learning & Stacked Autoencoders.

- What is "deep" learning? / Wide vs Deep / feature extraction and why it is important.
- Challenges of training deep networks & solutions: Layer-wise training / Memory using gates (LSTMs, GRUs)
- Stacked Autoencoders (SAEs): what they are / Why they are valuable / How to train them / How to use them for predictions. / feature extraction / When not to use them.

Lab 7: Introduce Caffe's "levels" and students will construct a 3-level SAE manually for dengue input.

Lecture 8: Autocaffe's Network Generator

- What it is / why it is valuable.
- Brief intro to Mini scripting.
- Using Prefabs

Lab 8: Implementing SAEs using Autocaffe's prefabs.

Homework: Retry the dengue prediction with a 5-layer SAE. Same time horizons. Students need to report their best scores on a special forum chat. We will select groups to present next day (10 mins + 10 mins Q&A). Everyone needs to prepare their presentation.

DAY 5

Lecture 9: Data Transformations

- Avoid dependencies in inputs. NNs are not black boxes
- Using Moments -- what are they / why they are useful for time series data / How to do this using Autocaffe.
- Thresholding
- Fourier transforms -- what it is / when is this useful?
- Optional Topic: Wavelets for noise reduction (we may add this topic in if time and

resources permit).

Lab 9: Retry dengue prediction with moments and thresholding of inputs. Read: the paper on earthquakes.

Lecture 10: Putting it All Together

- An efficient NN design and training workflow from start to finish – how to begin designing networks for real world problems?
- We will use the Dengue problem set to illustrate this workflow.

Homework: Implement Terra Weather's dengue network and see if you can improve results. Students need to report their best scores on a special forum chat. We select groups to present next day (10 mins + 10 mins Q&A). Everyone needs to prepare their presentation.

DAY 6

We will use Day 6 to wrap up and Q&A. However, we may have an optional Lecture.

Lecture 11: Ensembles (OPTIONAL)

- What are ensembles? / Why they are useful?
- Basic techniques
- Visualization -- Spaghetti Diagrams
- Implementing ensembles with Autocaffe.

Lab 11: Redoing the dengue problem with a 6-member ensemble (OPTIONAL).

0.4 Links

[1] http://caffe.berkeleyvision.org/
Caffe is a deep learning framework originally meant for machine vision. The emphasis is on speed. Terra Weather have adapted Caffe for timeseries forecasting.

[2] https://www.moh.gov.sg/content/moh_web/home/diseases_and_conditions/d/dengue.html
Dengue is a vector bourne disease, carried by the Aedes aegypti mosquito. This disease is strongly affected by the weather.

[3] https://www.virtualbox.org/wiki/Downloads
Please download and install version 5.2.0 on your laptop.

About Terra Weather

Terra Weather is a young technology company specializing in developing cutting-edge prediction technology and Artificial Intelligence based solutions. Our weather AI processes data from multiple sources such as satellites and sensors and convert them into highly accurate weather predictions. We enable multi-national Oil and Gas, Marine Transportation and Marine Consultancy companies work safely and successfully worldwide.

We are Hiring

We are actively looking for data scientists, engineers and programmers to join us solve interesting and groundbreaking applications in agriculture, maritime and energy sectors. Successful candidates can look forward to working on time-series prediction, distributed programming, and artificial intelligence technology. If you are passionate about AI and interested to join the team, please take the time to solve our programming challenge:

http://www.terra-weather.com/ly/challenge

and send your answer and CV to recruit@terra-weather.com.